Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 58 (2015) 524 – 529

Procedia Computer Science

Second International Symposium on Computer Vision and the Internet (VisionNet'15)

A Robust Algorithm for Speech Polarity Detection Using Epochs and Hilbert Phase Information

D. Govind, P. M. Hisham, D. Pravena

Center for Computational Engineering and Networking

Amrita Vishwa Vidyapeetham (University), Coimbatore, Tamilnadu

Abstract

The aim of the proposed work presented in this paper is to determine the speech polarity using the knowledge of epochs and the cosine phase information derived from the complex analytic representation of original speech signal. The work presented in this paper is motivated by the observation of variations in the cosine phase of speech around the Hilbert envelope (HE) peaks according to the polarity changes. As the HE peaks represent approximate epochs location, the phase analysis is performed by using algorithms which provide better resolution and accuracy of estimated epochs in the present work. In the present work, accurate epochs locations are initially estimated and significant HE peaks are only selected from the near vicinity of the epochs location for phase analysis. The cosine phase of the speech signal is then computed as the ratio of signal to the HE of speech. The trend in the cosine phase around the selected significant HE peaks are observed to be varying according to the speech polarity. The proposed polarity detection algorithm shows better results as compared with the state of the residual skewness based speech polarity detection (RESKEW) method. Thus, the improvement in the polarity detection rates confirms significant polarity information present in the excitation source characteristics around epochs location in speech. The polarity detection rates are also found to be less affected for different levels of noise addition which indicates the effectiveness of the approach against noises. Also, based on the analysis of mean execution time, the proposed polarity detection algorithm is confirmed to be 10 times faster than the RESKEW algorithm. [1].

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)

Keywords: Speech polarity detection, cosine phase, Hilbert envelope;

1. Introduction

The inversion in the polarity of the signals arises at the circuit level in the signal acquisition devices [1] [2]. For instance, the electrical connection fed to the inverting /non-inverting amplifiers of the data acquisition devices produce polarity

[1] The present work is archived at `arxiv.org` with identification number as 1407.3398(cs.SD)and titled as "Speech Polarity Detection Using Hilbert Phase Information". The work can be downloaded from the following link: `http://arxiv.org/pdf/1407.3398v2.pdf`. Authors here by declare that this archived work is not published in any other conferences or journals.

E-mail address: govinddmenon,hishamthangalpms,d.pravena@gmail.com

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)

doi:10.1016/j.procs.2015.08.068

inversion or phase change by π radians in the input signal [1,3]. Even if the polarity of the speech signals are inverted, change in polarity of the signals are perceptually indistinguishable [1,2]. For the case of speech signals and electroglottogram (EGG), the chances of possible polarity inversion is in the circuitry used in the audio pre-amplifiers. Most of the audio pre-amplifiers use combinations of inverting and non-inverting types of operational amplifiers in multiple stages for amplifying the signal captured using sensors. Choice of inverting/non-inverting type is dependent on the input impedance and gain parameters of the amplifier. The polarity inversion in the signal occurs when the non-inverting output of one stage of amplification is fed as input to a non-inverting stage of amplification [4]. The polarity inversion can be avoided in two ways. Firstly, by implementing a polarity detection circuit at the output stage of a preamplifier. Secondly, by devising an algorithm which detects polarity of the already acquired signal and correct the signal polarity back to the original default polarity. The polarity detection and correction at the circuit level in the pre-amplifier stage of audio acquisition device is proposed in [4]. At the algorithmic level, the polarity of the acquired signal has to be determined by means of signal processing and re-invert the signal back to the original signal polarity. In this paper, a polarity detection at the algorithmic level is proposed.

Even though polarity inversion is perceptually indistinguishable, the polarity changes adversely affect estimation of the speech parameters. For instance, the instantaneous F_0 values estimated from speech with a default (positive) polarity and inverted polarity (negative) shows variations [5]. Therefore an automatic polarity detection is essential as a part of speech processing to ensure the accurate estimation of speech parameters.

By understanding the significance of polarity detection many researchers have proposed algorithms for the polarity detection [1,2,3,6,7]. Popular approaches are listed below:

- Ding et al. [5]: Based on the spurious glottal wave computed from the linear prediction (LP) residual of speech
- Saratxaga et al. [7]: Phase cut and relative phase shifts (RPS) for polarity detection from speech
- Drugman et al. [3]: Phase shifts of the oscillating moments around the F_0s which are estimated locally
- Drugman [1](RESKEW): By measuring the statistical skewness between LP residual of speech and asymmetric version of LP residual
- Govind et al. [2]: Based on the sign of cosine phase corresponds to the locations of Hilbert envelope (HE) peaks of speech signals

Apart from these works, the fast estimation method of polarity of speech is proposed based on the integrated linear prediction residual by Abhiram et al. in [6]. The proposed fast method of polarity estimation also signifies the variations in the excitation source characteristics at the epochs location according to the polarity changes in speech.

The objective of the present work is to extract the polarity information by observing the cosine phase variations occurring around the epochs in speech signals. Epochs are referred to the instants in speech at which the excitation of the vocaltract is maximum [8]. The polarity is determined from the trend of the slopes in the cosine phase around the significant Hilbert envelope (HE) peaks nearest to the accurately hypothesized epochs. The work presented in the paper is organized as follows: The proposed algorithm of speech polarity detection using the knowledge of epochs and cosine phase of analytic signal is described in Section 2. The performance of the proposed algorithm and the comparison with the state of the art RESKEW methods are explained in Section 3 as experimental results. Finally, Section 4 summarizes the work with scope for future work.

2. Polarity Detection Using Epochs and Cosine Phase of Speech Signals

2.1. Cosine Phase Computation from Complex Analytic signal

The complex analytical signal is derived from the speech to compute the cosine phase. The analytic signal representation has speech as the real part and Hilbert transform of the speech as the imaginary part [9]. The Hilbert envelope (HE) is obtained as the magnitude of the analytic signal. The cosine phase is then computed as, $cos(\phi[n]) = \frac{s[n]}{\sqrt{s^2[n]+s_h^2[n]}}$, where $s[n]$ and $s_h[n]$ are the original speech signal and Hilbert transform of speech, respectively. Also, the quantity $\sqrt{s^2[n] + s_h^2[n]}$ is the termed as the HE of speech.

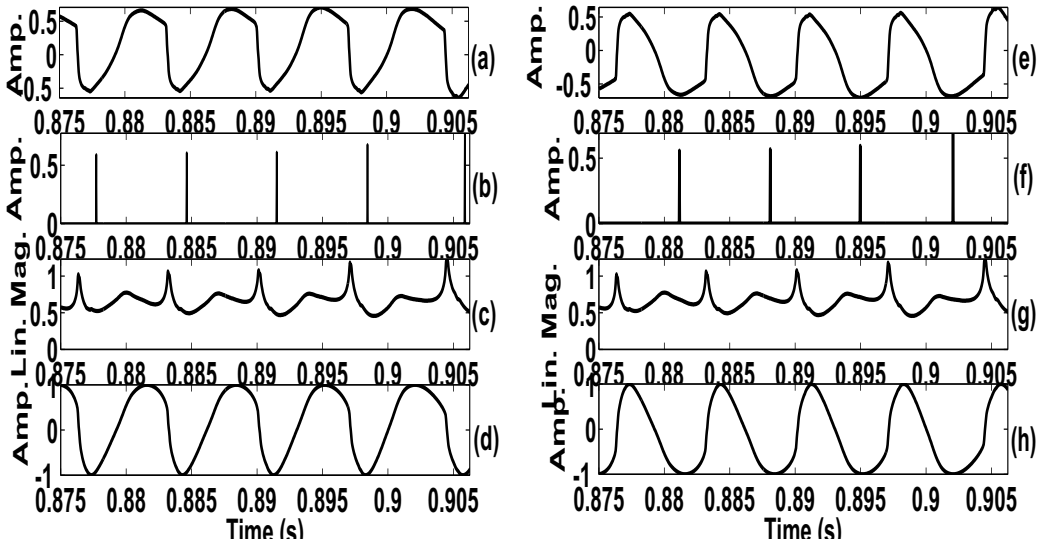

Fig. 1. Analysis of cosine phase around HE peaks in EGG. (a) original voiced segment of EGG , (b) reference epochs location obtained from (a), (c) HE of (a) and (d) its cosine phase segment. ((e)-(h)) show the equivalent plots obtained for polarity inverted of EGG of (a) which is obtained by multiplying all the EGG samples with -1.

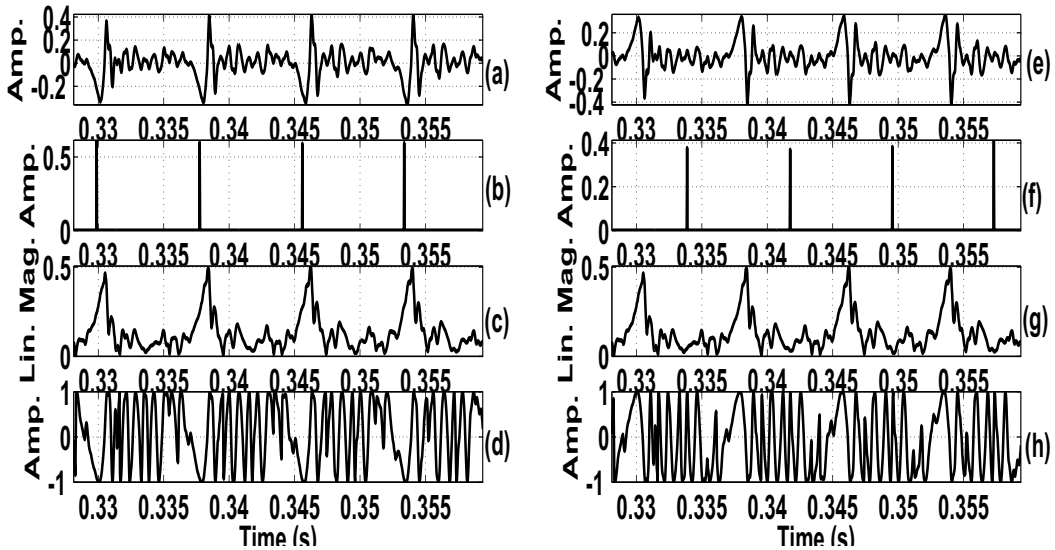

Fig. 2. Analysis of cosine phase around HE peaks in speech. (a) Voiced segment of original speech , (b) reference epochs location obtained from (a), (c) HE of (a) and (d) its cosine phase segment. ((e)-(h)) show the equivalent plots obtained for polarity inverted of voiced speech in (a).

Table 1. Performance comparison of proposed approach with RESKEW algorithm on CMU-Arctic and Berlin emotion speech (EmoDb) databases.

CMU Arctic Voice	Proposed Method			RESKEW Method		
	No. Correct	No. False	% Correct	No. Correct	No. False	% Correct
SLT	1130	0	100	1130	0	100
BDL	1130	0	100	1130	0	100
JMK	1112	0	100	1107	5	99.55
CLB	1131	0	100	1131	0	100
RSM	1131	0	100	1131	0	100
KSP	1131	0	100	1131	0	100
AWB	1131	0	100	1131	0	100
EmoDb	798	18	97.79	806	10	98.77
Total	8695	18	99.79	8697	15	99.8

Table 2. Performance comparison of proposed approach with RESKEW algorithm on CMU-Arctic and Berlin emotion speech (EmoDb) databases for simultaneously recorded EGG.

CMU Arctic Voice	Proposed Method			RESKEW Method		
	No. Correct	No. False	% Correct	No. Correct	No. False	% Correct
SLT	1130	0	100	1130	0	100
BDL	1130	0	100	1130	0	100
JMK	1113	0	100	1113	0	100
EmoDb	789	27	96.69	796	20	97.55
Total	4162	27	99.35	4169	20	99.52

2.2. Proposed Polarity Detection Using Epochs and Cosine Phase of Speech Signals

The polarity is determined by analyzing the trend in the cosine phase around locations of the HE peaks of speech signals. The locations of HE peak provide approximate locations of the epochs in speech. However, the presence of spurious HE peaks leads to erroneous cosine phase analysis which in turn result in the falsified decision about the polarity of the given speech. Hence the analysis of cosine phase segments should be anchored around relevant HE peaks. Hence a mechanism is essential to select the significant HE peaks. In the present work, the relevant HE peaks are selected by estimating the accurate locations epochs in speech. The relevant HE peaks are retained which are nearest to the epochs. The accurate epochs locations are extracted by the zero frequency filtering (ZFF) of speech signals as given by Murty et al.[8]. The trend in the zero crossings of the cosine phase is computed by retaining the corresponding HE peak locations nearest to each epoch . The trend in cosine phase computed as the slope at the zero crossings and is observed to vary according to the polarity of the signal.

The Figure 1((a)-(d))shows the voiced segment of EGG, hypothesized epochs, its HE and cosine phase having default positive polarity. Figure 1((e)-(h)) are obtained for the polarity reversal of Figure 1((a)). The cosine phase are observed in the regions around the HE peaks are plotted in Figure 1 ((c)-(d)). Also, the nature of the cosine phase is inverted when the polarity of the EGG is inverted as given in subplots ((d) & (h)) of Figure 1. Hence this trend in the cosine phase correspond to the HE peaks is considered as the primary evidences for the determining the polarity of the signal. Decision of the polarity is made by majority voting of number of positive or negative polarity detections obtained for each of the HE peaks in the signal. As a byproduct of the analysis, HE of EGG or speech is observed to be independent of the polarity of the signal (from Figure 1 ((c)&(g))). The polarity invariant nature of HE is utilized for reducing the ambiguities in locating the zero crossings in cosine phase corresponding to the HE peaks which arise due polarity reversal as time shifts in the estimated epochs (from Figure 1((b)&(f))). A similar trend occurs for the speech case also as plotted in Figure 2. Figure 2 demonstrates the effect of polarity in estimated epochs, HE and cosine phase in segment of voiced speech. The observed phase variations as observed in case EGG is also present in the speech and is plotted in Figure 2((d)&(h)) for both the polarity cases.

Table 3. Performance of speech polarity detection of the proposed method for additive babble noise for various SNRs.

SNR	0 (dB)	5 (dB)	10 (dB)	15 (dB)	20 (dB)	30 (dB)
Proposed	0.001%	0.000%	0.000 %	0.000%	0.000%	0.000 %
RESKEW	0.187%	0.087%	0.037 %	0.012%	0.0124%	0.024 %

Table 4. Performance analysis of average execution time (in Seconds) of CMU Arctic Voices.

Method	**CMU Arctic Voices**						
	SLT	BDL	JMK	KSP	CLB	AWB	RMS
Proposed	0.0471	0.0431	0.0428	0.0479	0.0454	0.0461	0.0482
RESKEW	0.4853	0.4741	0.4522	0.4722	0.4679	0.4703	0.4891

3. Experimental Results

The CMU-Arctic [10] and Berlin emotional speech databases [11] are used for the performance evaluation of the proposed method. The availability of the simultaneous speech and EGG recordings is reason for using the two databases for the performance evaluation. CMU-Arctic database has seven speakers and all speakers have recorded 1131 phonetically balance utterances. Out of the seven speakers three (BDL,JMK and SLT) speakers have both speech and EGG utterances. Each utterance in CMU-Arctic database is sampled at 32 kHz and with 16 bits resolution per sample. In contrast with the CMU-Arctic database, EmDb has emotional utterance of 10 German native speakers in seven emotional conditions in German language. Out of the seven emotions, four emotions (apart from neutral) are selected for study. Each utterance in EmDb is sampled at 16kHz and has a bit resolution 16 bits/sample. The performance of the proposed polarity detection algorithm is given in Table 1 and Table 2 for speech and EGG, respectively.

From the Table 1 and 2, the proposed polarity detection provides a near perfect detection from speech and EGG. Although the performance of the proposed method is inferior to that of the RESKEW method for EGG, the method provided improved polarity detection for clean speech signals.

From Table 3, polarity detection error rates of the proposed methods are not affected much for babble noise addition at the various SNR levels starting from 30 dB through 0 dB. The noise performances for both proposed and RESKEW methods are evaluated on the neutral speech signals EmoDb and CMU Arctic database. Hence utterances of all seven voices in CMU-Arctic database and a subset of EmoDb database excluding all the 712 emotion utterances, are used for the comparative evaluation of noise performances.

As the impulse like discontinuities at the epochs location are well preserved for noise addition with different noise levels in the speech, the polarity detection using the proposed method is least affected by the noise. The robustness against noise of proposed method is also observed to be similar to that of the RESKEW method.

3.1. Performance of Analysis of Execution Time

To compare the computational complexity experimentally, the execution time in determining the polarity for each utterance is computed for both proposed and RESKEW methods. For computing the execution time, the MATLAB commands 'tic' and 'toc' are used. For the time complexity analysis, the MATLAB code for implementing the proposed method is optimized by removing spurious loops and control statements. The MATLAB code available in GLOAT toolbox is used as it is for time complexity analysis of the RESKEW method. As the elapsed program execution time depends on the computer parameters, the programs for each of the two methods are executed in the same machine and on the same data set. After measuring the elapsed time taken in determining polarity of each utterance, the mean execution time is computed by finding the average across all the utterance available for each CMU speaker voice. Table 4 compares the mean execution time of proposed and RESKEW polarity detection algorithms for each CMU Arctic speaker voice. As there is a significant variation in the lengths of each utterance available for each emotion, EmoDb is not used for comparison. In contrast, each voice in the CMU-Arctic database provides almost equal number of phonetically balanced utterances with equal duration [10]. From the Table 4, each of CMU-Arctic speakers provide lower mean execution time for the proposed method as compared to RESKEW algorithm. Also from the Ta-

ble 4, the proposed polarity detection algorithm is observed to be nearly 10 times faster than the RESKEW algorithm.

4. Conclusions

An algorithm exploiting the cosine phase information anchored around the instants of significant excitation is developed for automatic determination of polarity of the speech signals. Based on the misclassification counts, the proposed algorithm showed better or equivalent result as compared to state of the art RESKEW method for the polarity detection of neutral speech signals under noise free conditions. The performance of the proposed method is also found to be least affected in presence of babble noise at various levels. As the instants estimation performance show degradation in emotive speech signals, the performance of the proposed polarity detection also degraded for speech in various emotions. The RESKEW method is found to be more robust for the polarity detection from emotion speech signals. Based on the execution time analysis, the proposed method is confirmed to be faster than the RESKEW algorithm. As compared to the existing cosine phase based method for polarity detection [2], the present work emphasizes the presence of additional polarity information around the epochs location in speech. In the present work, the spurious HE peaks of the speech signals are removed by considering the only peaks available in the approximate vicinity of epochs. Also, the block processing of the speech or EGG is avoided in the proposed algorithm by restricting the processing around epochs location.

References

1. T. Drugman, "Residual excitation skewness for automatic speech polarity detection," *IEEE Signal Process. Letters*, vol. 20, no. 4, pp. 387–390, April 2013.
2. D. Govind, A. S. Biju, and A. Smily, "Automatic speech polarity detection using phase information from complex analytic signal representations," in *Proc. SPCOM 2014*, 2014.
3. T. Drugman and T. Dutoit, "Oscillating statistical moments for speech polarity detection," in *Non-Linear Speech Processing Workshop (NOLISP11)*, 2011, pp. 48–54.
4. B. Deepak and D. Govind, "Significance of implementing polarity detection circuits in audio preamplifiers," 2015, accepted for publication in Int. Conf. on Advances in Computing, Communication and Informatics (ICACCI 2015).
5. W. Ding and N. Campbell, "Determining polarity of speech signals based on gradient of spurious glottal waveform," in *ICASSP 98*, 1998, pp. 857–859.
6. B. Abhiram, A. P. Prathosh, and A. G. Ramakrishnan, "A fast algorithm for speech polarity detection using long-term linear prediction," in *Proc. SPCOM 2014*. IISc Bangalore, 2014.
7. I. Saratxaga, D. Erro, I. Hernez, I. Sainz, and E. Navas, "Use of harmonic phase information for polarity detection in speech signals," in *INTERSPEECH 2009*, 2009, pp. 1075–1078.
8. K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Trans. Audio, Speech and Language Process.*, vol. 16, no. 8, pp. 1602–1614, Nov. 2008.
9. L. Cohen, *Time-Frequency Analysis: Theory and Applications*, S. P. Series, Ed. ser. Signal Processing Series. Englewood Cliffs: Prentice-Hall, 1995.
10. J. Kominek and A. Black, "CMU-Arctic speech databases," in *5th ISCA Speech Synthesis Workshop*, Pittsburgh, PA, 2004, pp. 223–224.
11. F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlemeier, and B. Weiss, "A database of German emotional speech," in *Proc. INTERSPEECH*, 2005, pp. 1517–1520.